1.68-angstrom crystal structure of endopolygalacturonase II from Aspergillus niger and identification of active site residues by site-directed mutagenesis

van Santen, Y; Benen, JAE; Schroter, KH; Kalk, KH; Armand, S; Visser, J.; Dijkstra, BW

Published in:
The Journal of Biological Chemistry

Publication date:
1999

1.68-Å Crystal Structure of Endopolygalacturonase II from Aspergillus niger and Identification of Active Site Residues by Site-directed Mutagenesis

Santen, Yovka van; Benen, Jacques A.E.; Schröter, Klaus-Hasso; Kalk, Kor H.; Armand, Sylvie; Visser, Jaap; Dijkstra, Bauke W.

Published in:
Default journal

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 274, No. 43, Issue of October 22, pp. 30474–30480, 1999

Printed in U.S.A.

1.68-Å Crystal Structure of Endopolygalacturonase II from *Aspergillus niger* and Identification of Active Site Residues by Site-directed Mutagenesis*

(Received for publication, July 19, 1999)

Yovka van Santen‡, Jacques A. E. Benen§, Klaus-Hasso Schröter‡¶, Kor H. Kalk‡, Sylvie Armand§, Jaap Visser§, and Bauke W. Dijkstra‡‖

From the ‡Laboratory of Biophysical Chemistry, Groningen University, 9747 AG Groningen, The Netherlands, and §Section on Molecular Genetics of Industrial Microorganisms, Wageningen Agricultural University, 6703 HA Wageningen, The Netherlands

Polygalacturonases specifically hydrolyze polygalacturonate, a major constituent of plant cell wall pectin. To understand the catalytic mechanism and substrate and product specificity of these enzymes, we have solved the x-ray structure of endopolygalacturonase II of *Aspergillus niger* and we have carried out site-directed mutagenesis studies. The enzyme folds into a right-handed parallel β-helix with 10 complete turns. The β-helix is composed of four parallel β-sheets, and has one very small α-helix near the N terminus, which shields the enzyme's hydrophobic core. Loop regions form a cleft on the exterior of the β-helix. Site-directed mutagenesis of Asp^{180}, Asp^{201}, Asp^{202}, His^{223}, Arg^{256}, and Lys^{258}, which are located in this cleft, results in a severe reduction of activity, demonstrating that these residues are important for substrate binding and/or catalysis. The juxtaposition of the catalytic residues differs from that normally encountered in inverting glycosyl hydrolases. A comparison of the endopolygalacturonase II active site with that of the P22 tailspike rhamnosidase suggests that Asp^{180} and Asp^{202} activate the attacking nucleophilic water molecule, while Asp^{201} protonates the glycosidic oxygen of the scissile bond.

The plant cell wall consists of a network of complex carbohydrates like cellulose, hemicellulose, and pectin. The latter is the most complex of these carbohydrates. It contains so-called "smooth regions" and "hairy regions." The smooth regions, also known as homogalacturonan, consist of α(1,4)-linked D-galacturonic acid residues, whereas the hairy regions, or rhamnogalacturonan I, are characterized by stretches of alternating α(1,2)-linked D-galacturonic acid and L-rhamnose (1). The rhamnose residues can be substituted at their O4 atoms by arabinose or galactose (2). Throughout the pectin molecule, the galacturonic acid residues can be methylated at O6 and/or acetylated at O2 and/or O3 (3). Due to its complex structure, modification of pectin by plants or complete breakdown by microorganisms requires many different enzymes.

In microorganisms several classes of pectinases have been identified. These classes comprise pectate-, pectin-, and rhamnogalacturonan lyases, rhamnogalacturonan hydrolases, and polygalacturonases, which all depolymerize the main chain; and pectin methylesterases and pectin- and rhamnogalacturonan acetylesterases, which act on the substituents of the main chain. Crystal structures are known of members of several classes of main chain depolymerizing pectinases. These include pectate lyases from *Erwinia chrysanthemi* and *Bacillus subtilis* (4–6), pectin lyases from *Aspergillus niger* (7, 8), and rhamnogalacturonase A from *Aspergillus aculeatus* (9). Recently, the crystal structure of endopolygalacturonase from the bacterium *Erwinia carotovora* was solved (10). The lyases cleave the substrate by β-elimination, whereas rhamnogalacturonases and polygalacturonases use acid/base-catalyzed hydrolysis (11, 12). Despite their completely different reaction mechanisms, and their groupings in different sequence homology families, the x-ray structures of pectate lyase, pectin lyase, and rhamnogalacturonase reveal a similar unique right-handed parallel β-helix topology (13, 14).

Together with the rhamnogalacturonases, the polygalacturonases have been assigned to family 28 of glycosyl hydrolases (15). Both endopolygalacturonase II and rhamnogalacturonase A act with inversion of configuration (16, 17), suggesting that the other family 28 glycosyl hydrolases also cleave their substrate with inversion of the anomeric configuration. The rhamnogalacturonases are specific for the strictly alternating α(1,4)-D-GalA-α(1,2)-L-Rha[1] sequence (11). In contrast, the polygalacturonases hydrolyze the α(1,4)-glycosidic bonds between adjacent galacturonic acid residues in the "smooth" part of the pectin molecule (12). However, the presence of a complete family of seven endopolygalacturonase encoding genes in *A. niger* (18) raises the intriguing question whether their gene products are all targeted to the same homogalacturonan part or whether some prefer other parts of the pectin molecule. Indications for the latter possibility have been recently described for endopolygalacturonase E from *A. niger* (19).

To understand the differences in action between the various endopolygalacturonases and to identify the residues that are critical for activity, we have elucidated the crystal structure of *A. niger* endopolygalacturonase II and have investigated the role of various residues in the active site by site-directed mutagene-

* This work was supported by European Union Project ERBBIO4CT960685. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 1czf) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

¶ Present address: AstraZeneca R & D, 34183 Mölndal, Sweden.

‖ To whom correspondence should be addressed: Laboratory of Biophysical Chemistry, Groningen University, Nijenborgh 4, 9747 AG Groningen, The Netherlands. Tel.: 31-50-363-4378; Fax: 31-50-363-4800; E-mail: bauke@chem.rug.nl.

[1] The abbreviations used are: Rha, rhamnose; GalA, galacturonic acid; PEG, polyethylene glycol; r.m.s.d., root mean square deviation; GlcNAc, *N*-acetylglucosamine; MES, 2-(*N*-morpholino)ethanesulfonic acid.

TABLE I
Data collection, MIRAS analysis, and refinement statistics for endopolygalacturonase II

	Native	Platinum[a]	Mercury[a]	Silver[a]
Data collection				
Soaking time		24 h	63 h	4 days
Resolution (Å)	1.68	2.53	2.35	2.53
a (Å)	65.50	66.51	66.28	66.31
b (Å)	201.24	203.32	203.27	202.42
c (Å)	49.07	49.65	49.61	49.67
No. of measured reflections	263,929	63,301	100,826	43,618
No. of unique reflections	69,332	21,165	28,220	16,031
Completeness (%)[b]	92.0 (64.9)	90.5 (51.6)	97.6 (65.9)	69.0 (11.5)
R_{sym}[c]	0.036 (0.078)	0.058 (0.069)	0.046 (0.077)	0.087 (0.060)[i]
MIRAS analysis				
No. of positions		4	11	4
R_{Cullis}[d]		0.525	0.501	0.676
Isomorphous phasing power[e]		1.70	2.19	0.91
Anomalous phasing power[e]		1.30	2.43	0.63
Overall figure of merit	0.681			
wARP				
No. of rounds	60			
w_{wARP}[f]	0.70			
Structure refinement				
Resolution (Å)	20–1.68			
No. of unique reflections	69,220			
No. of reflections in test set	3,575			
R_{crys}[g]	16.6%			
R_{free}[h]	19.1%			
r.m.s.d. bond lengths (Å)	0.007			
r.m.s.d. bond angles (°)	1.654			
Model	2 × 335 amino acid residues 2 GlcNAc 976 H_2O 6 Zn^{2+} ions			

[a] Platinum = 5 mM K_2PtCl_4; mercury = 5 mM HgC_6H_4COCHO; silver = saturated $AgCH_3COO$.
[b] Between brackets are the numbers for the highest resolution shell.
[c] $R_{sym} = \Sigma(|I - \langle I\rangle|)/\Sigma I$.
[d] $R_{Cullis} = \Sigma_{hkl}(||F_{PH} \pm F_P|) - F_H(calc)|/\Sigma_{hkl}|F_{PH} \pm F_P|$.
[e] Phasing power = $\Sigma_{hkl}|F_H(calc)|/\Sigma_{hkl}(|F_{PH}|) - (|F_{PH}(calc)|)|$.
[f] $w_{wARP} = F^2_{obs}(F^2_{obs} + \Sigma|F_{aver} - F_i|^2/n)$.
[g] $R_{crys} = \Sigma_{hkl}[|F_o(hkl, \text{work set})|] - [|F_c(hkl, \text{work set})|]/\Sigma_{hkl}[|F_o(hkl, \text{work set})|]$.
[h] $R_{free} = \Sigma_{hkl}[|F_o(hkl, \text{test set})|] - [|F_c(hkl, \text{test set})|]/\Sigma_{hkl}[|F_o(hkl, \text{test set})|]$.
[i] R_{sym}(highest resolution shell) is better than R_{sym}(overall) due to an ice ring in the diffraction pattern.

sis. These data enable us to propose for the first time a catalytic mechanism for family 28 glycosyl hydrolases.

EXPERIMENTAL PROCEDURES

Protein Purification, Crystallization, and X-ray Data Collection—The *pga*II gene was cloned and overexpressed, and the protein was purified as described (20). Upon secretion, the N-terminal 27 residues are cleaved off, resulting in the 35-kDa mature endopolygalacturonase II. This mature protein consists of 335 amino acids, numbered 28–362, and was shown to be heterogeneously glycosylated at Asn^{240} (21).

Previously, a crystallization protocol of endopolygalacturonase II was published in which the enzyme was crystallized from ammonium sulfate (22). However, that procedure very often resulted in the growth of twinned crystals. Much higher quality crystals could be obtained by crystallizing endopolygalacturonase II from PEG8000. We used sitting drop set-ups with a reservoir solution containing 100 mM sodium acetate, pH 6.0, 100 mM $ZnSO_4$, and 11–13% (w/v) PEG8000, and a drop containing a mixture of 2 μl of reservoir solution and 2 μl of protein solution (9 mg/ml in 10 mM sodium acetate, pH 6.0). Crystals of variable size, up to 0.2 × 0.4 × 0.5 mm^3, grew in 2–3 weeks. The crystals belong to space group $P2_12_12$ with cell dimensions a = 49.07 Å, b = 201.24 Å, c = 65.50 Å at 120 K, and two molecules per asymmetric unit. The Matthew's number, V_M, is 2.31 (assuming a molecular mass of 35 kDa and 8 molecules in the unit cell), indicating a solvent content of approximately 46.8%.

Prior to data collection and heavy atom soaking experiments, the crystals were soaked for at least 12 h in an artificial mother liquor containing 2.0 mM $ZnSO_4$, 15% (w/v) PEG8000, and 100 mM HEPES buffer, pH 7.5, for the native data set, the platinum derivative, and the mercury derivative, or 100 mM MES buffer, pH 6.5, for the silver derivative. Subsequently, crystals were soaked in artificial mother liquor containing the heavy atom compound (see Table I for details on soaking time and concentrations), whereafter the crystals were soaked for 15 min in a cryoprotectant (artificial mother liquor containing 25% v/v glycerol). This was directly followed by flash freezing of the crystal in a stream of evaporating nitrogen gas.

Native data were collected on the protein crystallography beam line at the ELLETRA Syncrotrone di Trieste, with a MAR image plate with λ = 1.0 Å. All derivative data sets were collected in house with a MacScience DIP-2030H image plate with CuK_α x-rays from a NONIUS FR591 rotating anode generator equipped with MacScience MAC-XOS double mirror focusing optics. Data were processed with DENZO and SCALEPACK (23).

Structure Determination—The structure of endopolygalacturonase II was determined by multiple isomorphous replacement with three heavy atom derivatives. The data of the crystal soaked in K_2PtCl_4 gave rise to a low (4 σ), yet consistent peak in a difference Patterson map. With phases calculated from this single platinum position, three more platinum positions were found in difference Fouriers. With the initial phases thus obtained, 11 heavy atom binding sites were established from difference Fouriers in the mercury derivative and 4 in the silver derivative. Refinement of the heavy atom parameters and phase calculations were carried out with the PHASES package (24). Both the isomorphous and the anomalous differences were included. The final MIRAS phases determined to 2.5-Å resolution had an overall figure of merit of 0.68. To improve these phases, solvent flattening was done with the DM (25) program from the CCP4 package (26). At this stage, a 2.5-Å electron density map was calculated and visually inspected with the program O (27). Although the map was judged to be interpretable, the phases were further improved and extended to 1.68 Å using the *wARP* procedure (28, 29) before any model building was attempted. After completion of the *wARP* procedure, model building was done with the aid of an automatic tracing procedure (30). In an iterative procedure the positions of dummy atoms from the best, lowest R-factor *wARP* dummy model were used in combination with the amino acid sequence to automatically trace the polypeptide chain, followed by improvement of the positions of the remaining (not yet assigned) dummy atoms with ARP (31). With this procedure 97% of the main chain and about 50% of

the side chains of the protein could be automatically interpreted. In the first molecule 331 of the 335 amino acid residues had their main chain and 45% of their side chains automatically fitted. A "frameshift" of one residue was observed for the first 60 amino acids, which was corrected manually. For the second molecule 322 amino acids had their main chain automatically built, with about 58% of their side chains fitted into the electron density. The model building was completed manually with the program O, making use of the two-fold non-crystallographic symmetry. In the second molecule two loops (10 and 6 amino acids) were initially left out due to unclear electron density.

The model thus obtained was refined using X-PLOR (32). To monitor the refinement procedure, 5% of the reflections were left out of the refinement as a test set. Refinement cycles were alternated with manual rebuilding sessions with the program O. The two missing loops in the second molecule were gradually built in. The two molecules are very similar, superimposing with an r.m.s.d. of 0.28 Å. Only a few side chains show different conformations (Lys^{44}, Thr^{59}, Lys^{124}, Asn^{207}, Asn^{232}, Ile^{260}, Ser^{261}, Ile^{267}, Glu^{292}, Lys^{299}, Glu^{312}, Lys^{349}, Lys^{354}, Val^{359}, and Ser^{361}). Most of these residues are located in loop regions. Water molecules were added to the structure with ARP. During the refinement procedure, clear electron density representing an *N*-acetylglucosamine residue appeared in both molecules. This GlcNAc was built in, connected to Asn^{240}. No electron density was observed for additional sugar units of this glycan. The geometry of the final model was checked with PROCHECK (33) (see also Table I).

RESULTS

Structure Determination of Endopolygalacturonase II—Details of the structure determination and the crystallographic refinement can be found in Table I. Endopolygalacturonase II crystallizes with two molecules in the asymmetric unit, called A and B, respectively. Superposition of the two molecules revealed that the 335 C_α atoms overlay with an r.m.s.d. of 0.28 Å. For the final model the r.m.s. coordinate error lies around 0.16 Å, as estimated from a Luzzati plot (34). The final electron density map was in general well defined. Two main chain regions in molecule B, comprising residues 228–236 and residues 292–296, have disordered electron density, suggesting conformational flexibility. In addition, for a few hydrophilic side chains pointing into the solvent, no clear electron density was obtained (for molecule A: Lys^{44}, Lys^{71}, Lys^{295}, and Lys^{349}; for molecule B: Lys^{44}, Lys^{71}, Lys^{124}, Lys^{299}, and Lys^{354}). The geometry of the model is good. In the Ramachandran plot, taking into account only the non-glycine and non-N- or -C-terminal residues, 484 residues (82.9%) are in the most favored region, 98 residues (16.8%) in the generously allowed regions, and 2 residues (0.3%) in the additionally allowed regions.

Overall Structure of Endopolygalacturonase II—Endopolygalacturonase II folds into a right-handed parallel β-helical structure comprising 10 complete turns (Fig. 1) with overall dimensions of approximately 65 Å × 35 Å × 35 Å. The number of amino acids per turn varies from 22 to 39, averaging to 29 residues per turn. This variation is caused by the diversity of lengths of the loops connecting the β-strands. The average rise per turn is 4.8 Å, a value typical for β-helical structures (13, 14). The β-helix is formed by four parallel β-sheets, named PB1, PB2a, PB2b, and PB3. This naming was adopted to be consistent with the naming of the β-sheets in the pectate lyase structure, the first right-handed parallel β-helical structure that was solved (4). In contrast to the three β-sheets in pectate lyase, the endopolygalacturonase II β-helix is composed of four β-sheets. PB1, PB2b, and PB3 are the endopolygalacturonase II counterparts of PB1, PB2, and PB3, respectively, of pectate lyase. The five-stranded PB2a sheet, located in β-helix turns 6 to 10, can be regarded as an N-terminal extension to sheet PB2b. Its strands are connected to those of PB2b via one residue in a left-handed α-helix conformation. This residue changes the direction of the main chain by approximately 100°. In pectate lyase, PB1 and PB2 are connected by turns, which are shorter than the strands of PB2a in endopolygalacturonase II. However, even though endopolygalacturonase II has an extra β-sheet, its overall shape is the same as that of pectate lyase. The β-helix is not perfectly cylindrical, as PB1 and PB2b are almost anti-parallel, with PB3 making an angle of approximately 95 degrees with PB2b (Fig. 1*b*). An overview of the strands in the parallel β-helix is presented in Table II.

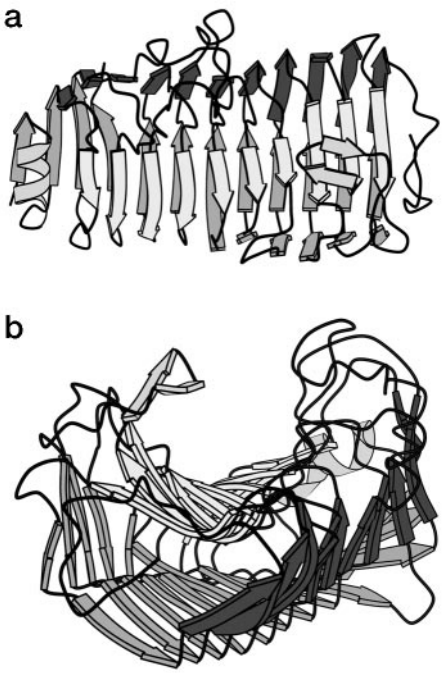

FIG. 1. *a*, the three-dimensional structure of endopolygalacturonase II with the N terminus on the *left* and the C terminus on the *right*, viewed onto β-sheet PB1 (*light gray*). PB2a and PB2b are shown in *gray*, and PB3 is shown in *dark gray*. *b,* the structure viewed from the C-terminal side, showing the cleft that is formed by the loop regions T1 (left side loop region) and T3 (right side loop region).

Besides the β-strands forming the β-helix, two other secondary structure elements are present in endopolygalacturonase II. One is a small two-stranded antiparallel β-sheet located in a loop between PB1 and PB2a (residues 290–292 and 295–296). The other is a small α-helix (residues 35–42) near the N terminus, between the first two strands of PB2a, which shields the hydrophobic core of the β-helix at the N-terminal side. Such an α-helical "cap" is observed in most β-helical structures (35). Additionally, the C-terminal side of the β-helix core is shielded from the solvent, by the C-terminal residues 353–362.

The turns between the β-helical strands are named based on the sheets they connect. The turns between PB1 and PB2 (a or b) are referred to as T1-turns, between PB2 (a or b) and PB3 as T2-turns, and between PB3 and PB1 as T3-turns. The T2 turns are the most regular ones. They are short, consisting of one or two amino acid residues, making a smooth connection between PB2b and PB3. The only exception is the T2 turn between β8 and β9 (see Table II). It contains four amino acids (residues 102–105), which bulge out of the β-helix. The T1- and the T3-turns are more diverse. T1-turns comprise 2–15 amino acids (7, 4, 2, 3, 3, 2, 9, 5, 15, and 5 amino acids for the consecutive turns, respectively) and the size of the T3-turns varies from 2 to 20 amino acids (5, 9, 20, 6, 15, 5, 4, 3, 2, and 4 amino acids for the consecutive turns, respectively). The T1-turns are relatively longer near the C-terminal side of the β-helix, whereas the T3-turns are longest near the N-terminal side of the β-helix. In this way the loops form two bulky extensions on the exterior of the β-helix (Fig. 1b). Between these extensions a large cleft is present, the bottom of which is formed by PB1. The cleft is approximately 8 Å wide, and well suited to accommo-

TABLE II
Assignment of β-strands in the right-handed parallel β-helix of A. niger endopolygalacturonase II

Turn	β-Sheet					
	PB1	PB2a	α_L[a]	PB2b	α_L[a]	PB3
				β1 29–32		
				β2 47–51		β3 54–55
1	β4 61–63			β5 71–75		β6 77–80
2	β7 90–93			β8 98–101		β9 106–108
3	β10 129–134			β11 137–140		β12 143–145
4	β13 152–155			β14 159–162		β15 165–167
5	β16 182–184			β17 188–193		β18 196–198
6	β19 204–206	β20 209–210	N211	β21 212–215		β22 218–220
7	β23 225–227	β24 237–239	N240	β25 241–244	H245	β26 246–249
8	β27 253–259	β28 265–268	E269	β29 270–273	N274	β30 275–280
9	β31 283–288	β34 304–307	D308	β35 309–312	S313	β36 314–318
10	β37 323–328	β38 334–335	D336	β39 337–340	D341	β40 342–345

[a] α_L, residues that are in a left-handed α conformation, which connect two adjacent β-strands, are shown in one-letter code with their sequence number. Additional secondary structure elements, not shown in this table, are one α-helix (residues 35–42) and one two-stranded anti-parallel β-sheet (β32, residues 290–292, and β33, residues 295–296).

date the unbranched polygalacturonan substrate. It is open on both sides, in accordance with the endohydrolytic character of the enzyme. Moreover, several residues located at the bottom of the cleft are indispensable for substrate binding and/or catalysis (see below), indicating the functional importance of the cleft.

Endopolygalacturonase II contains four disulfide bridges, which are strictly conserved in all *A. niger* endopolygalacturonases. Two disulfide bridges, one in the N- and one in the C-terminal region (Cys^{30}-Cys^{45} and Cys^{353}-Cys^{362}, respectively), ensure the "capping" of the core of the β-helix. The N-terminal disulfide bridge forces the α-helix of residues 35–42 to fold over the N-terminal side of the β-helix. The C-terminal disulfide bridge pulls residues 360–362 over the C-terminal end of the β-helix. The third disulfide bridge, Cys^{203}-Cys^{219}, connects two adjacent β-helical turns in the middle of the putative active site cleft. Finally, the Cys^{329}-Cys^{334} disulfide bridge connects the beginning and the end of the T1 loop in turn 10, keeping the end of PB1 and the beginning of PB2a together.

Glycosylation and Ion Binding Sites—A. niger endopolygalacturonase II contains a single glycosylation site at Asn^{240} (21). In both molecules in the asymmetric unit extra electron density was found extending from this asparagine into which one *N*-acetylglucosamine residue could be built. No density was visible for additional carbohydrate residues. The *N*-acetylglucosamine residue points into the solvent region and is not involved in crystal contacts. It is located on the exterior of the β-helix, on the opposite side of the cleft where the substrate is supposed to bind.

In addition, two zinc ions were located in the asymmetric unit. Zinc ions are absolutely essential for crystallization of the enzyme from PEG8000, and crystals dissolve if no zinc is present in the mother liquor. In both the A and B molecules, one zinc ion mediates crystal contacts between Asp^{110} from one molecule and Asp^{308} and Asp^{336} from a symmetry-related one. A water molecule is the fourth ligand of each zinc ion.

Four other electron density peaks were interpreted as zinc ions as well, two in each molecule. One is bound to His^{223} and is octahedrally coordinated by its Nϵ nitrogen atom and 5 water molecules. The other one is involved in crystal contacts, bound octahedrally to Nϵ of His^{96}, the main chain N, and the side chain Oδ atoms of Asp^{28} from the non-crystallographic symmetry-related molecule and three water molecules. They have six ligands at a distance of around 2.2 Å. The coordination number, the distance to the ligands, and the height of the electron density peaks are not compatible with the presence of a water molecule at these positions. To confirm the nature of these ions, an anomalous difference Fourier was calculated with data collected at the zinc edge (λ = 1.31 Å) (collected at the EMBL X31 beam line at the DORIS storage ring, DESY, Hamburg, Germany). Apart from a 20-σ peak at the position of the Zn^{2+} ion mediating the crystal contacts near Asp^{110}, Asp^{308} and Asp^{336}, a 10-σ peak showed up near His^{223}, but no peak was present near His^{96}. However, the anomalous data were collected at pH 6.0, the original pH of crystallization, and at that pH histidine side chains may be (partly) protonated reducing the affinity for positively charged ions. This may explain the lower electron density of the peak near His^{223}, and the absence of density near His^{96}. Indeed, a preliminary refinement of the model against the pH 6.0 data set showed that the occupancy of the ion near His^{223} refined to about 0.7, that the density for the putative ion near His^{96} had disappeared, and that one of the other ligands of this ion, Asp^{28}, had become disordered. Asp^{28} is well defined at pH 7.5, the pH of the high resolution native structure. Therefore, we conclude that in the high resolution native structure the electron density peaks near His^{223} and His^{96} are most likely Zn^{2+} ions. In addition, a survey of zinc binding sites in proteins and small molecule compounds (36) revealed that zinc ions can have an octahedral coordination sphere with ligands at 2.1–2.2 Å.

Exploring the Active Site by Site-directed Mutagenesis—Comparison of the available polygalacturonase sequences from bacterial, fungal, and plant origin reveals that only eight amino acid residues are strictly conserved (37). These residues are: Asn^{178}, Asp^{180}, Asp^{201}, Asp^{202}, His^{223}, Gly^{224}, Arg^{256}, and Lys^{258} (sequence numbering according to *A. niger* endopolygalacturonase II). The eight conserved residues form a predominantly negatively charged patch in the cleft. Space to accommodate a sugar substrate is available above the plane formed by Asp^{180}, Asp^{201}, Asp^{202}, and His^{223}. Gly^{224} is buried in the cleft. It has φ/ψ angles not allowed for other amino acid residues (φ is 99° and ψ is −167°), and no space for a side chain is available (Fig. 2).

To assess their importance for activity, residues Asp^{180}, Asp^{201}, Asp^{202}, His^{223}, Arg^{256}, and Lys^{258} were mutagenized. Replacement of His^{223} by Ala resulted in an enzyme with only 0.5% of wild type activity and no effect on the K_m for the substrate (38). Replacement of Asp^{180}, Asp^{201}, and Asp^{202} by Asn and Glu resulted in mutant enzymes that had remaining specific activities compared with wild type of 0.01/0.08%, 0.01/0.01%, and 0.6/0.01% for D180E/N, D201E/N, and D202E/N, respectively. The K_m values had only changed minimally. In contrast, the R256N and K258N enzymes showed not only a reduced specific activity (14% and 0.8% of wild type specific activity, respectively), but also an approximately 10-fold de-

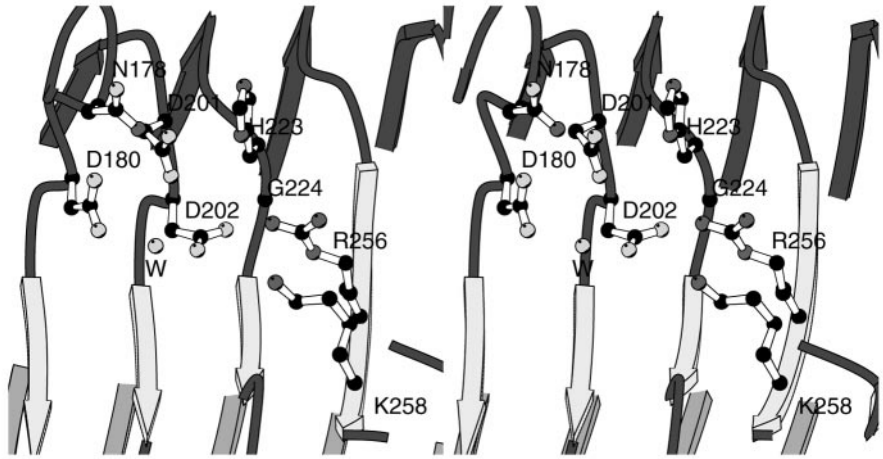

FIG. 2. **A stereo view of the active site cleft, looking onto PB1.** Residues that are completely conserved among all polygalacturonases, Asn^{178}, Asp^{180}, Asp^{201}, Asp^{202}, His^{223}, Gly^{224}, Arg^{256}, and Lys ^{258}, are shown in ball-and-stick. The putative hydrolytic water molecule is indicated by *W*.

crease in K_m. This shows that Asp^{180}, Asp^{201}, Asp^{202}, and His^{223} might be directly involved in catalysis and that Arg^{256} and Lys^{258} might play a role in substrate binding.

DISCUSSION

Comparison to Other Right-handed Parallel β-Helix Proteins—The first enzyme found to contain the right-handed parallel β-helix motif was pectate lyase C (4). Since then, several other structures with β-helical topology have been determined. They comprise several pectin degrading enzymes, namely pectin and pectate lyases, rhamnogalacturonase, and polygalacturonase (4–10). In addition, right-handed parallel β-helix structures have been observed for *Bordetella pertussis* virulence factor P.69 pertactin (39), a protein involved in polysaccharide recognition, and the phage P22 tailspike protein (40, 41). This latter protein has rhamnosidase activity.

Of these enzymes, endopolygalacturonase from *E. carotovora* and rhamnopolygalacturonase A are the most similar to *A. niger* endopolygalacturonase II, in agreement with their classification into the same glycosylhydrolase homology family, family 28 (15). The *A. niger* and *E. carotovora* endopolygalacturonases have 10 complete turns; the rhamnopolygalacturonase has one additional turn on the C-terminal side. All have four β-sheets that form the β-helix, and their loop regions are also similar, in both size and location. In the core of right-handed parallel β-helix structures, four types of stabilizing side chain-side chain interactions have been recognized: aliphatic stacking interactions, aromatic stacking interactions, asparagine ladders, and serine stacks (13, 14). The number of interactions and the interaction types are variable. The pectate lyases contain all four interaction types, whereas in the polygalacturonases only aliphatic stacking and aromatic stacking interactions occur. In the endopolygalacturonase II core, aliphatic stacking interactions are predominant (Ile^{48}-Val^{72}-Val^{99}-Ile^{139}; Val^{190}-Ile^{212}-Val^{241}-Ile^{270}-Val^{309}; Ile^{107}-Ile^{144}-Ile^{166}-Ile^{197}; Val^{255}-Val^{285}-Ile^{325}; Val^{154}-Val^{184}-Val^{206}-Ile^{227}-Ile^{257}-Ile^{187}-Leu^{327}; Val^{238}-Val^{267}-Ile^{306}); only one case of aromatic stacking is observed (Phe^{129}-Phe^{152}-Phe^{182}). In contrast, two occurrences of threonine stacking are present on the surface of endopolygalacturonase II (Thr^{242}-Thr^{271}; Thr^{140}-Thr^{162}).

Comparison of Family 28 Glycosyl Hydrolases—The *E. carotovora* endopolygalacturonase is the only other polygalacturonase with known structure (10). This bacterial enzyme shows only 19% sequence identity to the *A. niger* endopolygalacturonase II. Nevertheless, the two structures are very similar, with 265 equivalent C_α atoms (out of 335) superposable with an r.m.s.d. of 1.8 Å. The β-helix strands display the highest similarity. *E. carotovora* endopolygalacturonase has, however, large insertions in the T3 loops of β-helix turns 1 and 2. These insertions increase the size of one side of the putative active site cleft in *E. carotovora* endopolygalacturonase. In contrast, the other side of the cleft, formed by the T1 loops, is similar in size despite some insertions and deletions in the various T1-turns. Notwithstanding the differences in the T1 and T3 loop regions, the width and direction of the cleft are approximately the same in both enzymes. Further differences between the two enzymes are located at the terminal sides of the β-helix. Insertions in the N-terminal region make the *E. carotovora* endopolygalacturonase β-helix wider, and a long N-terminal tail, which is absent in *A. niger* endopolygalacturonase II, folds along the exterior of its β-helix. *E. carotovora* endopolygalacturonase lacks the C-terminal residues 360–362 that shield the C-terminal end of the core of the β-helix. In this respect the *A. niger* polygalacturonase II more resembles the *A. aculeatus* rhamnogalacturonase than the *E. carotovora* polygalacturonase. Additionally, the fungal enzymes contain more structurally aligned residues and the four disulfide bridges are conserved among them, whereas they are not conserved in the *E. carotovora* endopolygalacturonase. This may indicate that the evolutionary divergence of the fungal β-helical proteins from the bacterial ones occurred before the divergence of the polygalacturonases and rhamnogalacturonases.

The Catalytic Mechanism—Two often observed catalytic mechanisms for glycosyl hydrolases involve two acidic residues (42, 43). In the case of retaining enzymes these residues are involved in a double displacement mechanism, via a covalent glycosyl-enzyme intermediate and their carboxylate groups are spaced at approximately 5.5 Å. In inverting enzymes, the distance between the side chains of the acidic residues is approximately 9.5 Å and catalysis proceeds via a single displacement mechanism. In this mechanism one of the acidic residues acts as a general acid, donating a proton to the glycosidic oxygen of the scissile bond. The second carboxylate acts as a general base, which activates a water molecule that performs a nucleophilic attack on the sugar anomeric carbon.

Endopolygalacturonase II is an inverting enzyme (16). Surprisingly, the distances between the absolutely conserved aspartates (average of the Oδ1-Oδ1, the Oδ1-Oδ2, the Oδ2-Oδ1, and the Oδ2-Oδ2 distances) are 4.1 Å (Asp^{180}–Asp^{201}), 5.7 Å (Asp^{180}–Asp^{202}) and 4.9 Å (Asp^{201}–Asp^{202}), respectively, and also the conserved histidine residue (His^{223}) is close by (3.5 Å from Asp^{201}, 4.0 Å from Asp^{180}, and 5.0 Å from Asp^{202}, respectively). No conserved carboxylic acid residues are found at a distance of about 9.5 Å from each other. This suggests that the

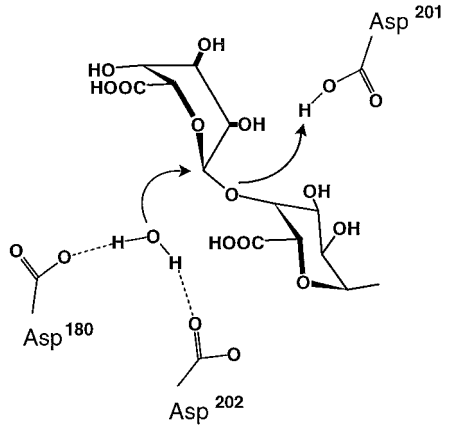

FIG. 3. **Schematic representation of the catalytic mechanism proposed for family 28 glycosyl hydrolases.** The conserved Asp^{201} (*A. niger* endopolygalacturonase numbering) acts as the proton donor, while Asp^{180} and Asp^{202} activate the hydrolytic water molecule.

family 28 enzymes do not conform to the "standard" inverting mechanism.

Further insight into the catalytic mechanism was obtained by a comparison of the *A. niger* polygalacturonase structure with that of the phage 22 tailspike rhamnosidase (40). This protein, which has not been classified into a glycosyl hydrolase family, resembles the β-helix fold of the family 28 enzymes. It has also three acidic residues in the active site (Glu^{359}, Asp^{392}, and Asp^{395}). Carbohydrate binding experiments with this protein suggested catalytic functions for these three residues (41, 44). A water molecule is bound between Glu^{359} and Asp^{395} in a position suitable for direct nucleophilic attack of the C1 carbon atom of the scissile glycosidic bond. The Asp^{392} side chain is at hydrogen bonding distance from the O1 atom of the -1 sugar (see Davies *et al.* (45) for binding site nomenclature), suggesting that it may be the proton donor in the reaction (41, 44).

In *A. niger* polygalacturonase II, a similar arrangement of acidic active site residues exists. A water molecule is bound between Asp^{180} and Asp^{202}, while the Asp^{201} carboxylic group is in a somewhat less solvent-exposed environment stabilizing its protonated state. This suggests that in endopolygalacturonase II Asp^{201} could be the proton donor and Asp^{180} and Asp^{202} could activate the water molecule (schematically shown in Fig. 3). A superposition of endopolygalacturonase II and the P22 tailspike protein with bound substrate, based on the positions of the four catalytically important oxygen atoms, shows that a galacturonate residue can be modeled in an orientation resembling that of the rhamnose residue in the tailspike protein. At the reducing and non-reducing ends of this saccharide, sufficient space is available to accommodate a longer saccharide chain without clashes with protein atoms. In this orientation the oligosaccharide is bound with its reducing end in the direction of the C terminus of the protein.

Experimental evidence for this mode of substrate binding comes from a site-directed D282K mutant.[2] In the proposed orientation of the substrate, residue 282 would be part of subsite $+2$. The mutant retained 60% of wild type activity, but its product bond cleavage frequencies on penta- and hexagalacturonides had changed. Compared with wild-type enzyme, more galacturonate monomers were produced and fewer dimeric oligosaccharides. This is consistent with Asp^{282} being part of subsite $+2$ and the existence of four subsites binding the non-reducing end of the oligosaccharide.

Further support for the proposed direction of substrate binding was obtained from a comparison of polygalacturonase II and rhamnogalacturonase. While polygalacturonases cleave the $\alpha(1,4)$-glycosidic bond between two D-galacturonate residues, rhamnogalacturonases hydrolyze $\alpha(1,2)$-linkages between a D-galacturonate and L-rhamnose. For both enzymes the newly formed reducing end sugar is a galacturonate residue. Thus, for both enzymes the -1 subsite is expected to be similar, while the $+1$ site will be different. Of the eight absolutely conserved residues in the polygalacturonases, only four are also present in *A. aculeatus* rhamnogalacturonase A: Asp^{180}, Asp^{201}, Gly^{224}, and Lys^{258}, respectively. Furthermore, Asp^{202} is a glutamate in rhamnogalacturonase but its side chain occupies approximately the same position as the side chain of Asp^{202}. Therefore, Asp^{180}, Asp^{201}, and Asp^{202}, and possibly Lys^{258}, are most likely part of subsite -1. On the other hand, Arg^{256}, which is conserved in the polygalacturonases only, probably constitutes subsite $+1$. This would agree with our proposed direction of substrate binding.

The mechanism described above provides no obvious role for the catalytic residue His^{223}. However, the -1 rhamnose residue in the tailspike protein was observed in a distorted boat conformation, and modeling of an undistorted galacturonate trimer (46) in the active site of endopolygalacturonase II leads to clashes between the enzyme, close to His^{223}, and the galacturonate residue in the $+1$ subsite. Moreover, upon substrate binding, breaking of the hydrogen bond between the O2 of the galacturonate moiety in subsite $+1$ and one of the carboxylate group atoms of the residue in subsite -1 would facilitate departure of the leaving group. His^{223} may play a role in the breaking of this hydrogen bond. Alternatively, it may aid in the distortion of the galacturonate residue in subsite -1. Further research is required to unambiguously establish the precise role of His^{223} in catalysis.

In summary, our approach of x-ray crystallography combined with site-directed mutagenesis of *A. niger* endopolygalacturonase II has revealed for the first time a possible catalytic mechanism for family 28 glycosyl hydrolases. Asp^{201} is proposed to act as the acid (proton donor), while Asp^{180} and Asp^{202} activate the hydrolytic water molecule. Despite an arrangement of active site residues completely different from that normally observed in inverting glycosyl hydrolases (43), the concerted action of an acid and a base has been conserved.

Acknowledgments— We are very grateful to Anastassis Perrakis (EMBL, Grenoble, France) for carrying out the automatic tracing of endopolygalacturonase II. The staff of the protein crystallography beam lines at the EMBL outstation in Hamburg (supported through European Union Large Installations Project Contract CHGE-CT93-0040) and at the ELLETRA Sincrotrone di Trieste are acknowledged for their help with data collection.

REFERENCES

1. de Vries, J. A., Rombouts, F. M., Voragen, A. G. J., and Pilnik, W. (1982) *Carbohydr. Polym.* **2,** 25–33
2. O'Neill, M. A., Albersheim, P., and Darvill, A. G. (1990) in *Methods in Plant Biochemistry:Carbohydrates* (Dey, P. M., ed) Vol. 2, pp. 415–441, Academic Press, London
3. Schols, H. (1995) *Structural Characterization of Pectic Hairy Regions Isolated from Apple Cell Walls.* Ph.D. thesis, Agricultural University, Wageningen, The Netherlands
4. Yoder, M. D., Keen, N. T., and Jurnak, F. (1993) *Science* **260,** 1503–1507
5. Lietzke, S. E., Scavetta, R. D., Yoder, M. D., and Jurnak, F. (1996) *Plant Physiol.* **111,** 73–92
6. Pickersgill, R., Jenkins, J., Harris, G., Nasser, W., and Robert-Baudoy, J. (1994) *Nat. Struct. Biol.* **1,** 717–723
7. Mayans, O., Scott, M., Connerton, I., Gravesen, T., Benen, J., Visser, J., Pickersgill, R., and Jenkins, J. (1997) *Structure* **5,** 677–689
8. Vitali, J., Schick, B., Kester, H. C. M., Visser, J., and Jurnak, F. (1998) *Plant Physiol.* **116,** 69–80
9. Petersen, T. N., Kauppinen, S., and Larsen, S. (1997) *Structure* **5,** 533–544
10. Pickersgill, R., Smith, D., Worboys, K., and Jenkins, J. (1998) *J. Biol. Chem.* **273,** 24660–24664
11. Suykerbuyk, M. E. G., Kester, H. C. M., Schaap, P. J., Stam, H., Musters, W., and Visser, J. (1997) *Appl. Environm. Microbiol.* **63,** 2507–2515
12. Kester, H. C. M., and Visser, J. (1990) *Biotech. Appl. Biochem.* **12,** 150–160

[2] J. A. E. Benen, unpublished results.

13. Jurnak, F., Yoder, M. D., Pickersgill, R., and Jenkins, J. (1994) *Curr. Opin. Struct. Biol.* **4,** 802–806
14. Yoder, M. D., and Jurnak, F. (1995) *FASEB J.* **9,** 335–342
15. Henrissat, B. (1991) *Biochem. J.* **280,** 309–316
16. Biely, P., Benen, J. A. E., Heinrichová, K., Kester, H. C. M., and Visser, J. (1996) *FEBS Lett.* **382,** 249–255
17. Pitson, S. M., Mutter, M., van den Broek, L. A. M., Voragen, A. G. J., and Beldman, G. (1998) *Biochem. Biophys. Res. Commun.* **242,** 552–559
18. Bussink, H. J. D., Buxton, F. P., Fraaye, B. A., de Graaf, L. H., and Visser, J. (1992) *Eur. J. Biochem.* **208,** 83–90
19. Parenicová, L., Benen, J. A. E., Kester, H. C. M., and Visser, J. (1998) *Eur. J. Biochem.* **251,** 72–80
20. Bussink, H. J. D., Kester, H. C. M., and Visser, J. (1990) *FEBS Lett.* **273,** 127–130
21. Yang, Y., Bergmann, C., Benen, J., and Orlando, R. (1997) *Rapid Commun. Mass Spectrom.* **11,** 1257–1262
22. Schröter, K.-H., Arkema, A., Kester, H. C. M., Visser, J., and Dijkstra, B. W. (1994) *J. Mol. Biol.* **243,** 351–352
23. Otwinowski, Z. (1993) *Proceedings of the CCP 4 Study Weekend: Data Collection and Processing,* Daresbury Laboratory, Warrington, UK
24. Furey, W., and Swaminathan, S. (1998) *Methods Enzymol.* **277,** 590–620
25. Cowtan, K. (1994) *Joint CCP 4 ESF-EACBM Newsl. Protein Crystallogr.* **31,** 34–38
26. Collaborative Computational Project No. 4 (1994) *Acta Crystallogr. Sec. D* **50,** 760–767
27. Jones, T. A., Zou, J.-T., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. Sec. A* **47,** 110–119
28. Perrakis, A., Sixma, T. K., Wilson, K. S., and Lamzin, V. S. (1997) *Acta Crystallogr. Sec. D* **53,** 448–455
29. van Asselt, E. J., Perrakis, A., Kalk, K. H., Lamzin, V. S., and Dijkstra, B. W. (1998) *Acta Crystallogr. Sec. D* **54,** 58–73
30. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) *Nat. Struct. Biol.* **6,** 458–463
31. Lamzin, V. S., and Wilson, K. S. (1993) *Acta Crystallogr. Sec. D* **49,** 129–147
32. Brünger, A. T., Kuriyan, J., and Karplus, M. (1987) *Science* **235,** 458–460
33. Morris, A. L., MacArthur, M. W., and Thornton, J. M. (1992) *Proteins* **12,** 345–364
34. Luzzati, V. (1952) *Acta Crystallogr.* **5,** 802–810
35. Jenkins, J., Mayans, O., and Pickersgill, R. (1998) *J. Struct. Biol.* **122,** 236–246
36. Alberts, I. L., Nadassy, K., and Wodak, S. J. (1998) *Protein Sci.* **7,** 1700–1716
37. Kester, H. C. M., Kusters-van Someren, M. A., Müller, Y., and Visser, J. (1996) *Eur. J. Biochem.* **240,** 738–746
38. Benen, J. A. E., Kester, H. C. M., Parenicová, L., and Visser, J. (1996) *Prog. Bio/Technol.* **14,** 221–230
39. Emsley, P., Charles, I. G., Fairweather, N. F., and Isaacs, N. W. (1996) *Nature* **381,** 90–92
40. Steinbacher, S., Seckler, R., Miller, S., Steipe, B., Huber, R., and Reinemer, P. (1994) *Science* **265,** 383–386
41. Steinbacher, S., Miller, S., Baxa, U., Budisa, N., Weintraub, A., Seckler, R., and Huber, R. (1997) *J. Mol. Biol.* **267,** 865–880
42. McCarter, J. D., and Withers, S. G. (1994) *Curr. Opin. Struct. Biol.* **4,** 885–892
43. Davies, G., and Henrissat, B. (1995) *Structure* **3,** 853–859
44. Steinbacher, S., Baxa, U., Miller, S., Weintraub, A., Seckler, R., and Huber, R. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 10584–10588
45. Davies, G., J, Wilson, K., S, and Henrissat, B. (1997) *Biochem. J.* **321,** 557–559
46. Walkinshaw, M. D., and Arnott, S. (1981) *J. Mol. Biol.* **153,** 1055–1073